A Semi-Automatic Framework to Discover Epistemic Modalities in Scientific Articles

Sviatlana Danilava
JW Goethe-University Frankfurt am Main
Dept. of Computer Science and Mathematics
Robert-Mayer-Str. 11-15, D-60486 Frankfurt am Main, Germany.
Email: danilava@cs.uni-frankfurt.de

Christoph Schommer
University of Luxembourg
Dept. of Computer Science - ILIAS Laboratory, MINE Research Group
6, Rue Richard Coudenhove-Kalergi, 1359 Luxembourg, Luxembourg
Email: christoph.schommer @uni.lu Home: mine.uni.lu

April 8, 2008

Abstract

Documents in scientific newspapers are often marked by attitudes and opinions of the author and/or other persons, who contribute with objective and subjective statements and arguments as well. In this respect, the attitude is often accomplished by a linguistic modality. As in languages like english, french and german, the modality is expressed by special verbs like can, must, may, etc. and the subjunctive mood, an occurrence of modalities often induces that these verbs take over the role of *modality*. This is not correct as it is proven that modality is the instrument of the whole sentence where both the adverbs, modal particles, punctuation marks, and the intonation of a sentence contribute. Often, a combination of all these instruments are necessary to express a modality. In this work, we concern with the finding of modal verbs in scientific texts as a pre-step towards the discovery of the attitude of an author. Whereas the input will be an arbitrary text, the output consists of zones representing modalities.

1 Introduction

Search engines that base on the World-wide Web find large amounts of hits and information by any request. However, intelligent search queries like Which scientists support hypothesis A or Does the author believe in my opinion are not yet supported. In order to answer, a search engine must first search for appropriate documents and then analyse them fast. For this, intelligent algorithms are required that take into account linguistic insights for a analytical consideration of syntax and style but also for a treatment with meta-aspects like opinion and attitude of the author himself. This is especially true for scientific texts: they are very objective in concern of the description of a hypothesis or in discussing diverse problems. The discovery of the subjective opinion or attitude of the author is a major topic and is the research objective of *Attitude Mining*. It concerns with the discovery of meta-information out of documents, especially the attitude of an author in respect to events, references to other's work, etc. The attitude can be positive or negative, but in most cases, it is hidden and to be proved by indications. Attitude Mining concerns with the explorative discovery of these indications ([21]), but demands for profound knowledge in areas like computer science, linguistics, cognitive sciences and psychology. Following [8], there exist more than 350 lexical style attributes for the attitude, for example to express doubts or beliefs. To further motivate, the following sentences should demonstrate the existence of subjectivity in scientific texts:

→ When paleontologists seek the roots of life, they head to rocks of the Archaean Eon, which range from 3.8 billion to 2.5 billion years old.

→ Australian and Canadian researchers argue this week in Nature that stromatolites were so diverse and complex that they must have been alive.

→ Martin Brasier of Oxford University is less sanguine, arguing that the structures are more likely chemical precipitates. He also objects to the reasoning in the Nature paper. "You cant use the argument that complexity is the signature for life," he says.

The first sentence is neutral, as it describes only a procedure what palaeontologists normally do *when they try out to find the origin of life*. The second sentence holds a hypothesis with explanatory statements, the third sentence arguments against the hypothesis in the second sentence having the author/originator referenced. In this respect, the modality concerns with the speaker's style to modify the proposition of sentences through subjective

components. And as we have seen above, many sentences are modal, for example

→ I believe she arrives this morning at London Heathrow.

→ I can not be in today.

In western languages, the modality is expressed by special verbs like can, must, may, etc. and by the subjunctive mood. However, this often induces that verbs take over the role of *modality*, which is not correct: it is proven that modality is the attribute of the whole sentence where both the adverbs, modal particles, punctuation marks, and the intonation of a sentence contribute to it. Often, a combination of all these instruments are necessary to express a certain modality. For example, the sentence

→ Do you really think that?

leads to another understanding as with

→ You do not really think of that?

In the first sentence, the combination of *really*, *think* and the transfer into a question is very subjective, but leaves the recipient some space. However, the second sentence is much more subjective, influencing the recipient's answer completely and leaving no space for another answer than 'no'. Overall, the complexity in using modalities is one of the major problems, both for the analysis of texts per se and for machine translation systems.
In this work, we concern with finding modal verbs in scientific texts as a prestep toward discovering the attitude of an author. Whereas the input will be an arbitrary text, the output consists of zones representing modalities.

2 Fundamentals

Originally, the concept of modality derives from the formal logic. Here, a modal expression consists of two parts, the *modal* part and the *proposition* part. The modal part contains the modality, the proposition the actual statement. Moreover, the modal part is either *deontic* or *epistemic* ([18]). A deontic modality describes the conditions that leads the statement to true or false, always being in relation with the reality, for example:

→ Indeed, the turnover of phytoplankton can be so high that there can be inverted pyramids of biomass, in which the standing crop of herbivorous zooplankton actually exceeds that of the phytoplankton.

The verb acts as modality, it expresses a certain idea of the objective reality that might come true under certain circumstances. The epistemic modality, on the other side, concerns with personal experiences and a knowledge level of the author, but less with reality:

→ Australian and Canadian researchers argue this week in Nature that stromatolites were so diverse and complex that they must have been alive.

The verb *must* appears in an epistemic way, the statement is not proven yet but still an assumption. This assumption is proven initiated by a justification. Furthermore, the source of information is given, for example in

→ Martin Brasier of Oxford University is less sanguine, arguing that the structures are more likely chemical precipitates.

This sentence contains an explicit source, namely *Martin Brasier*. Such statements are referenced as evidential statements and are mostly referenced as a sub-category of an epistemic modality.
The modality is supported by a set of expressions: in order to develop a methodology in respect to an automatic recognition, the lexical fundament must be found first. Modal verbs form a class of verbs that add a modal meaning to a proposition. They allow the sender to modify the essence of a sentence by possibilities, necessities, doubts, beliefs, etc. In the English language, this is for example must - have to, can - could - may, and will - would - shall. However, the use of modal verbs often leads to ambiguity as the same modal verbs are taken to express both the deontic and the epistemic relevance. Verbs like believe, doubt, accept, reject, etc. describe the mental state of the speaker or his attitude against propositional part of the statement. Moreover, *noun* may describe the mental states or cognitive processes as well, for example by *doubt, belief, rejection, etc.*. Adverbs and adjective are *lexical modifiers* that may assign doubts and beliefs, for example *perhaps, probably, possibly, certain, likely*, etc.
English modal verbs are used both in epistemic and in deontic meanings. Generally, modal verbs express either a possibility or a necessity; each modal verb offers several meanings with semantic and pragmatic differences, for example the word *must*. In the deontic version, it describes a necessity with the consideration of an external source, where the propositional subject is

not source of modality. In contrast to this, an epistemic version describes a necessity, taking a logical justification. The following two sentences are deontic (first) and epistemic (second):

→ I must go, she is already waiting for me.

→ Where is John? It is 14h00, he must be in school.

The epistemic reading of modal verbs can be summarised as follows:

- Epistemic necessity as a conclusion out of the speaker's evidence: *she must be in her office.*
- Epistemic necessity as logical conclusion out of a common valid and known fact: *she will be in her office.*
- Epistemic possibility as an uncertainty of the speaker: *she may be in her office.*

The epistemic usage of modal verbs, the epistemic adverbs and cognitive verbs distribute the subjectivity. The provide a basis for the attitude of the author, as for example in

→ The individual grains in them could not have accumulated mechanically because the slope of the cone is too great, says Stanley Awramik, a stromatolite expert at the University of California, Santa Barbara, who was not involved in the research.

Here, the proposition is just a personal attitude (*could*) of the referenced person, that is not proven at all. Given by the modal verbs, there is still enough information to discover the author's attitude and to differentiate the author's attitude against others' attitudes.

3 Selected Research Work

The current research follows divergent directions, especially in the establishment of linguistic and cognitive models. These models support an understanding of the lexical means of expression, their influence to the lexical environment, and the modification of meaning while using modality.
In respect to modalities as a influencing component to discovering the attitude, [19] says that it is insufficient to implement the attitude as to be

positive or negative. Moreover, the attitude can be modified via *contextual valence shifters* by not, never, none, but must take into account modifiers like rather, deeply, and/or few. [1] says that a *reported speech* shares a particular attention, since evidential aspects must be examined additionally. [9] argues that the lexical means of expression should not become considered as conveyor of meaning, but typical structures of attitude phrases can be observed.
The analysis of lexical resources that is additionally used to highlight the intention of the authors to produce attitudes is currently under research as well. [15] follows an establishment of specific emotional lexicons with positive, negative, and neutral meaning as well as an automatic extraction of emotion to extend these lexicons.
The detection of document zones to structure the document becomes more and more popular. Initially, it has been presented as *Argumentative Zoning* by Teufel and Moens ([24]), but has been applied in other works as well ([22], [23]) or strongly influenced research work on *Content Zoning* ([5]). The main motivation is to *summarise* documents and to zone in discourse-rhetoric zones. Teufel and Moens argue that - depending on the type, genre and style of the text - a standardised structure can often be identified. Using *scientific articles*, they have assigned seven argumentative zones to each text, the zoning is then performed by a supervised learning system. [17] suggests an extend classification where each sentence is assigned to a rhetoric role. There exist up to ten zones that are classified into 3 classes. They argue that there exist no sequences of rhetoric roles; sentence may belong to different zones, also called as combined zones.
Following the idea of *Opinion Mining*, [4] describe a model to detect *opinion words*. The idea is to discover propositions, which contain subjective lexical expressions and the proposition itself, for example in combination with *accuse*, *criticise*, or *doubt*. All constituents of each sentence receive a zoning label like *Opinion Proposition*, *Opinion Holder* or *Null*. Another approach are disambiguation processes of modal verbs, where [12] has implemented a rule-based system towards the disambiguation of the epistemic and deontic meaning of the german verbs like sollen, können, or dürfen.

4 Architecture

The framework of this work consists of two major parts which are presented in Figure 1. The first part focus on pre-processing the input text whereas the second part concerns with the disambiguation of the modal verbs and semi-

automatic classification of the corresponding sentences. The pre-processing begins with a part-of-speech tagger, and is followed by a module to detect the naming entities and the pronouns.

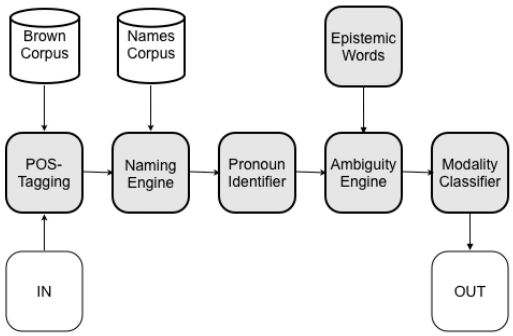

Figure 1: The architecture of the framework, using the Brown Corpus and the Names Corpus. First, a part-of-speech tagger[2] to a given input and sends the intermediate result to the *naming* and *pronoun engine*. After that, modalities are disambiguated (*Ambiguity Engine*) while synchronised with the list of modality verbs and finally classified (Modality Classifier). The text output is contains *modality tags*.

We have taken an advantage in a way that we have used the WordNet thesaurus ([7]) to establish a list of modality verbs. This list contains several lexical categories and are found recursively by using a synonym function of WortNet. Currently, there exist several corpora for the English language, for example the Brown University Corpus ([14]), the International Corpus of English (ICE), and the British National Corpus (BNC). The ICE is a set of corpora that supports various dialects of English from around the world. The BNC is a text corpus with both written and spoken English words, covering covers more than 100 million words of the late twentieth century from a wide variety of genres. However, in this work, we use the Brown University Corpus to support the part-of-speech tagger to assign a syntactic category to each word of the input document. Here, the word is kept as it occurs, meaning that the original word is substituted by a list of syntactic categories and the original word. Words of the same root but of another flexion are kept as they are.
In concern of dissolving *naming entities*, a first method concerns with identification of personal names on the basis of references that are probably given in the document. Per definitionem, this method is firstly applied but suffers

from diverse proper names of institution names like Max-Planck Institute. In this case, external databases must be consulted using an automaton (see Figure 2). For the identification of person names, we have used the *Names Corpus* by [13], which contains 5001 female and 3000 male first names.
To identify the pronouns in the text, we restrict the list of possible candidates and consider only *he*, *she*, and *who* as they concretely reference to one specific person. Common terms like *researchers* or *community members* are not considered as well as the pronoun *they* and *cataphora*. To identify the pronouns, we firstly concern with *who*, which occurs after a referenced nominal phrase (NP) but in the same sentence as a NP.
After having pre-processed the data, the annotated texts are then sent to the classification module. The modal verbs are first disambiguated before they are sent to the classifier. As we must differentiate between deontic and epistemic modality, these two classes are taken as classes. We then use the following simple rule scheme:

- A sentence is **deontic modal** if it contains a modality word that is deontic and if there exist modality words, which reference to facts.
- A sentence is **epistemic modal** if it contains a modality word that is epistemic and there exist modality words, which reference to subjective attitude of the author.
- A sentence is **non-modal** if there is no lexical evidence for modality.

The scheme may become improved when other criteria for disambiguation are included, for example the time. A more granular differentiation between *epistemic positive* and *epistemic negative* is possible when considering together the modal and propositional part of the sentence and classifying the sentences into *Author X believes in Y* (positive) and *Author rejects Y* (negative). The disambiguation process is shown as disambiguation automaton in Figure 3.
The automaton decides to which class a modal verb belongs to. Depending on certain collocations, the a-priori probability for a modal verb to be epistemic is generally higher than to be deontic, so that we take a decision quite early. For example, if a certain collocation proves that a modal verb v_i is probably epistemic for 90 percent, the automaton classifies v_i as to be epistemic. The classification criteria are:

- The modal verb must refers to an epistemic necessity if it occurs with the following components: have been, be and verb present participle,

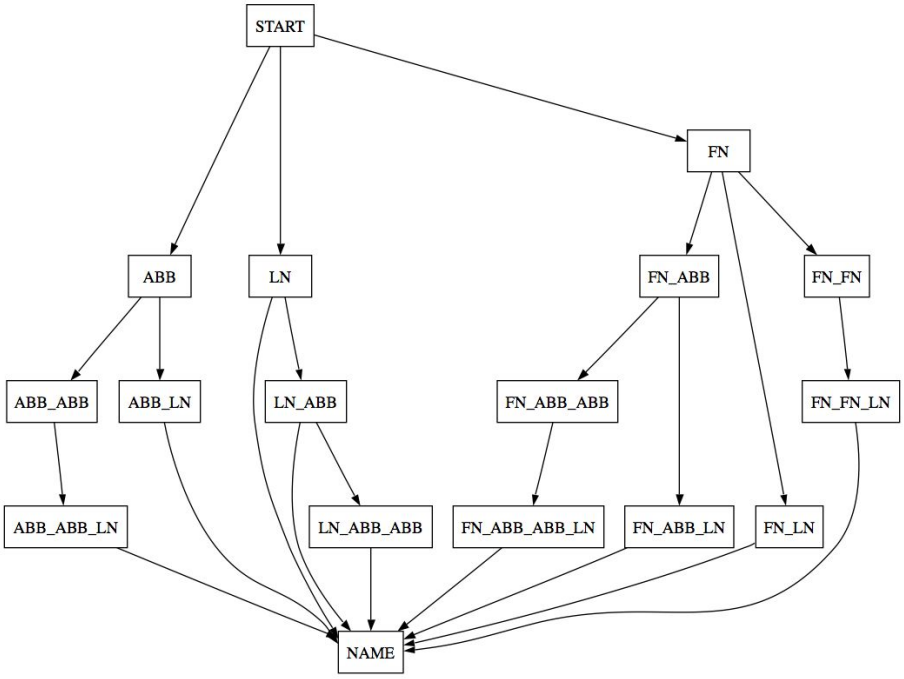

Figure 2: Automaton used for Naming Detection where *FN* corresponds to the full first name, *LN* to the full last name, and *ABB* to any kind of abbreviations, like the abbreviated middle name. For example, P. Green follows the path of *ABB_LN*, whereas *Peter Green* empties in *FN_LN*.

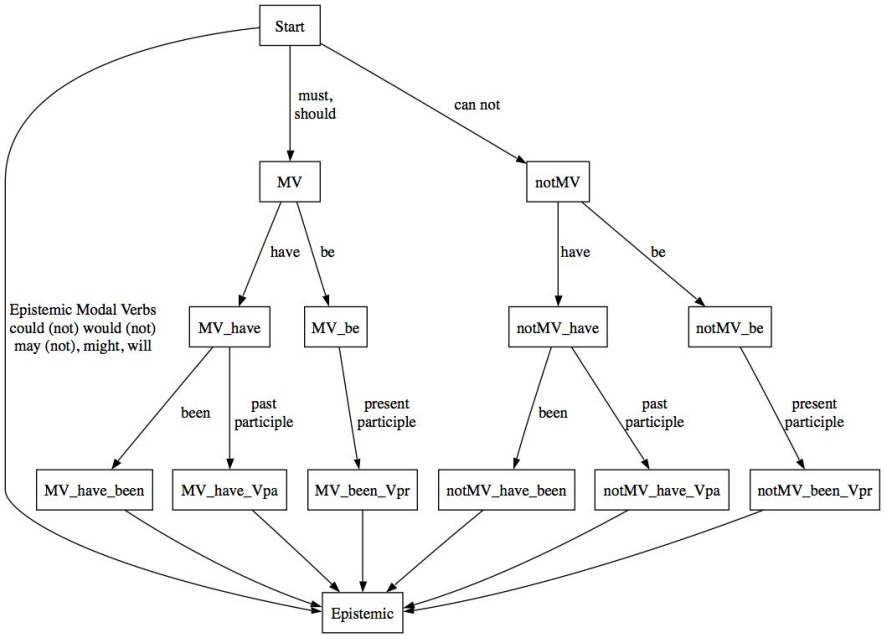

Figure 3: Automaton used for disambiguation where *MV* corresponds to modal verb, *Vpa* the verb past participle, and *Vpr* the verb present participle. *not* represents the negation, *have*, *be*, and *been* the corresponding words.

have and verb past participle, have been and verb present participle. In all other cases, the verb should be deontic.

- can is deontic.
- can not refers to the same verb components than must.
- could is epistemic.
- may is epistemic.
- might is epistemic as a tentative version of may.
- will is epistemic since future aspects are still hypothetic.
- shall is deontic.
- should is epistemic as must.

In Figure 3, only the paths to the class epistemic are shown; it is assumed that all other paths are either deontic or non-modal. A path like

MV→have→been

means that the sentence contains a sequence of verb and *have* and *been*. Modal verbs express the attitude and opinion of a person. In this work, we concern with two types of persons:

- The person is the author: the author of the text gives an opinion and attitude about hypotheses, other authors, or other methods. This often occurs in scientific articles, for example references or citations. In the following example, the attitude is expressed by the author himself:

  → Would a 100 mm scanning resolution be sufficient to produce an accurate model for paleontological study, or is a 50 mm scanning resolution a requirement?

- The person is the third person: this happens if the author speaks about other persons and presents those attitudes. In this case, these persons are referenced explicitly by name or work. This is a typical way of discussions in scientific articles.

  → Lowe pointed out their resemblance to modern forms but later had doubts.

To estimate the attitude of an author, we only consider epistemic sentences. We mark the importance of the modal part by a predicate M and $\neg M$, respectively, and the propositional part by a predicate H, if the propositional part contains arguments pro M, otherwise $\neg H$. All epistemic sentences can be described with

$$M(H) \text{ or } M(\neg H) \text{ or } \neg M(H) \text{ or } \neg M(\neg H)$$

However, this step can become conditionally automated as it is quite hard to decide if the modal part is M or $\neg M$: to do so, we certainly must find out the lexical information about a modal verb inside its lexical environment. Some modifiers like less or more and negations like not or none must be taken into account as they modify the meaning of the modal verbs. Their scope is important; an exact analysis implies the definition of complex grammars. Secondly, we must decide if the propositional part is H or $\neg H$, so that we concern with propositional content analyis. This could be done with thesauri like WordNet, as these contain descriptions of relationships between words, for example synonyms. For example, WordNet allows a multiple calculation of similarity between words, depending on the distance between these words in the thesaurus: the shorter the distance, the similar the words.
In this work, we have identified two problems: first, the similarity between two words does not correspond to the actual situation in the text and second, the similarity can only be computed between pairs of words, but not between phrases or sub-phrases. We may say at this point that the architecture is *hybrid*, meaning that the last step of estimating the attitude is done manually - based on the result that is produced. We then finally get a text result that is composed of text and meta information, consisting of two parts: the first part is machine readable as the data structures stay constantly with tags and structural information; it can therefore further be processed. The second part contains the epistemic sentences. A third and last step concerns with the segmentation of epistemic sentences depending on the hypothesis of the text. For this, we may use a graph, where all referenced persons are classified into three classes: *Pro* references all members P, *Contra* all members C, and *Neutral* all members N. Each group can be empty, but not at the same time, as the author must belong to at least one class. We then assign

- *Pro* refers to sentences of $M(H)$ and $\neg M(\neg H)$
- *Contra* to sentences of $M(\neg H)$ and $\neg M(H)$

- *Neutral* collects undecidable sentences, especially of those persons who decline a decision.

5 Example

The following steps show an example using the following scientific text:

→ "The individual grains in them could not have accumulated mechanically because the slope of the cone is too great," says Stanley Awramik, a stromatolite expert at the University of California, Santa Barbara, who was not involved in the research.

Generally, figures, formulas, and charts are manually pre-processed and substituted by tag-placeholders like *FIG* or *MATH*. The part-of-speech tagger then marks the text by two subsequent loops, where first all words are matched against the *Brown Corpus*. Often, domain-specific termini arise, which are unknown and therefore labeled by a *None*. Therefore, a second loop takes into account the morphologic structure of these words, for example, assigning a suffix *tion* to the category *noun*:

→ [(The, ART), (individual, ADJ), (grains, NNS), (in, IN), (them, PPO), (could, MV), (not, *), (have, HAVE), (accumulated, VPA), (mechanically, RB),...]

Recognizing the names, we then check if the text contains a list of references: in the positive case, all names are marked by a *Person*-tag. These words that begin with an uppercase letter are considered as well and set to candidates of possible first and last names, abbreviations, or other personal names. They are marked by a *NP*-tag. The first names are matched up with the mentioned *Names Corpus*. However, as ambiguity may occur, such words are disambiguated manually. We then get the following automaton as it has been described in Figure 3:

→ ... <Person> (Stanley, NP) (Awramik, NP) </Person>, ...

The decision, to which objets a personal pronoun belongs to, is taken by considering the lexical categories *PPS* and *WPS*:

→ ... <Person> (Stanley, NP) (Awramik, NP) </Person>, ...
... <Person Name= Awramik> (who, WPS) </Person Name= Awramik>

The final classification then leads us to

→ <EPISTEMIC>
... (could, MV), (not, *), (have, HAVE), (accumulated, VPA), ...
</EPISTEMIC>

where the tag *EPISTEMIC*, *DEONTIC*, or *NON-MODAL* represent the modal state. As modelled in Figure 3, the phrase could not → have → accumulated is ambiguous and leads to *negMV_HAVE_VPA*. The sentence therefore is marked as epistemic.

6 Classification Results

We have used scientific articles from the fields of palaeontology and biology as a first test set (in the following called SCA) and contributions to the scientific newspaper (in the following called SCI) as a second test set. All test documents of SCA share a common frame like *Author X talks about his work Y*; text documents of SCI share a frame like *Author X talks about the opinions of M scientists in respect to hypothesis Y*. For SCA, the texts share a similar length and style; the number of epistemic sentences is dominant to deontic and/or non-modal sentences (see Figure 4).

Text	Reports Total	Reports Epistemic (%)	Reports Deontic (%)	Reports Non-Modal (%)
1	67	40.9	7,5	51.6
2	61	31.1	21.3	47.6
3	90	20	6.7	73.3
4	39	20.5	0	79.5
5	71	12.7	11.3	76.0
6	60	36.7	8.3	55.0
7	87	29.9	6.9	63.2
8	73	20.5	6.8	72.7
9	63	33.3	6.3	60.4
10	41	19.5	0	81.5

Text	Reports Total	Reports Epistemic (%)	Reports Deontic (%)	Reports Non-Modal (%)
1	36	19.4	0	80.6
2	84	17.9	2.4	20.3
3	131	32.8	10.7	56.5
4	33	33.3	3.0	63.7
5	54	7.4	7.4	85.2
6	71	31.0	2.8	66.2
7	28	28.6	7.1	64.3
8	37	21.6	8.1	70.3
9	36	11.1	2.8	86.1
10	34	50	5.9	44.1

Figure 4: Percental distribution of epistemic, deontic, and non-modal sentences, where the left chart corresponds to SCA, the right chart to SCI.

Text	SCA Total	SCA correct (%)	SCA wrong (%)	SCI Total	SCI correct (%)	SCI wrong (%)
1	27	66.6	33.4	7	85.7	14.3
2	19	68.4	31.6	15	73.3	26.7
3	18	88.9	11.1	43	86.0	14.0
4	8	75.0	25	11	100	0
5	9	100	0	4	75	25
6	22	77.3	22.7	22	95.5	4.5
7	26	76.4	23.6	8	87.5	12.5
8	15	86.8	13.2	8	75.0	25.0
9	21	76.2	23.8	4	75.0	25.0
10	8	100	0	17	94.1	5.9
TOTAL	173	78.6	21.4	139	87.0	13.0

Figure 5: Percental classification result of selected sentences of SCA and SCI. The correct classified sentences are higher for SCI (87%) than to SCA (78.6%).

In total, 312 sentences have been analysed where 55.4% are of SCA and 44.6% from SCI. As presented in Figure 5, the correct classified sentences for SCI (87%) are higher than to SCA (78.6%). In respect to the wrong classified sentences, the modal word will occurs most frequently. The following list shows some epistemic sentences that are classified correctly and wrongly:

→ EPISTEMIC This evidence of an ecological shift preceding phenotypic change suggests that this part of the sequence **may** record rapid evolution driven by shifts in trophic ecology and adaptation to benthic niches.(correct)

→ EPISTEMIC If this **hypothesis** is correct however the low number of specimens displaying intermediate phenotypes is puzzling and the scenario of replacement of one lineage by another cannot be ruled out. (correct)

→ EPISTEMIC Yet direct evidence that feeding controls evolution over extended time scales available only from the fossil record is difficult to obtain because it is rarely **possible** to directly analyze dietary change in long-dead animals. (wrong)

→ EPISTEMIC First **perhaps** the best-known work on specialisation in fishes concerns stickleback in postglacial coastal lakes in Canada where

planktivores and benthic feeders coexist as two reproductively isolated and phenotypical distinct tropic.(wrong)

→ EPISTEMIC Laboratory feeding experiments and analyses of wild stickleback populations **show** that micro-wear exhibits a progressive shift from planktivores to benthic feeders.(wrong)

The main reason for a wrong classification is that - although the modal verb only has influenced a part of the whole sentence - the whole sentence has been assigned as to be epistemic. Especially, composed sentences like

→ EPISTEMIC-DEONTIC This uncertainty **may** relate to the fact that Buddenbrockia genes have undergone rapid sequence evolution, which **can** either cause artifactual groupings or reduce the support for the correct grouping.

have been classified twice, i.e., being epistemic and deontic. This is wrong as only the first part (may) is epistemic, the second part deontic (can).

7 Conclusions

The calculation process is characterised and influences by a multitude of external contributions, therefore, one of the next steps will be a step-by-step automatisation and the access to extended sources.
Although the classification result show good results, a more detailed consideration of modal verbs may become concerned as some of them negatively and positively influence propositional sentences. Last, the lexical environment must be considered if we want to automate the general hypothesis of being the modal part is M or $\neg M$. If a modal verb is discovered in the sentence structure, we can assume that the meaning is either positive or negative; it can be modified, if negations occur.
We still have in mind to constitute the modality as one possible method to characterise the author's attitude. This may be accomplished by other works of the group, i.e., the zoning of textual documents, the imaging of texts to self-organizing maps, and the fingerprinting of texts using statistic and linguistic variables.

8 Acknowledgement

This work has been performed within the research project *TRIAS*, which is funded by the University of Luxembourg.

References

[1] S. Bergler: Conveying attitude with reported speech. In Computing Attitude and Affect in Text: Theories and Applications, pp. 1122, 2006.

[2] S. Bird, E. Loper, and E. Klein. The Natural Language Toolkit. Version 9.0. 2007.

[3] A. L. Berger, S. Della Pietra, and V. J. Della Pietra: A maximum entropy approach to natural language processing. Computational Linguistics, 22(1):3971, 1996.

[4] S. Bethard, H. Yu, A. Thornton, V. Hazivassiloglou, and D. Jurafsky: Extracting opinion propositions and opinion holders using syntactic and lexical cues. In Computing Attitude and Affect in Text: Theories and Applications. pp. 125141. Springer, 2006.

[5] C. Brucks, M. Hilker, C. Schommer, C. Wagner, and R. Weires: Semi-automated Content Zoning of Spam Emails. Lecture Notes on Business Information Processing (Springer).

[6] S. Danilova: Semi-automatische Bestimmung der Attitüde über epistemische Modalität. Diplomarbeit. JW Goethe-University, Frankfurt am Main. Feburary 2008.

[7] C. Fellbaum. Wordnet: An Electronic Lexical Database. Bradford Books, 1998.

[8] J. Holmes: Doubt and certainty in esl textbooks. Applied Linguistics, 9, 1. pp. 2044, 1988.

[9] J. Karlgren, G. Eriksson, and K. Franzen: Where attitudinal expressions get their attitude. In Computing Attitude and Affect in Text: Theories and Applications, pages 2331, 2006.

[10] B. Kipper: Eine Disabiguierungskomponente für Modalverben. In KONVENS, pages 258267, 1992.

[11] B. Kipper: MODALYS - a system for the semantic-pragmatic analysis of modal verbs. In AIMSA, pp. 171180, 1992.

[12] B. Kipper. Ambiguitätsprobleme bei der Modalverbanalyse, 1995.

[13] M. Kantrowitz, B. Ross. Names corpus. Carnegie Mellon, 1991.

[14] H. Kucera, W. N. Francis, Brown University. 1967.

[15] Y. Y. Mathieu: A computational semantic lexicon of french verbs of emotion. In Computing Attitude and Affect in Text: Theories and Applications. pp. 109124, 2006.

[16] R. Mitkov: Anaphora resolution: The state of the art, 1999.

[17] Y. Mizuta, T. Mullen, and N. Collier. Annotation of biomedical texts for zone analysis. Technical report, National Institute of Informatics, 2-1-2 Hitotsubashi, Chiyoda, Tokyo, Japan, 2004.

[18] F. R. Palmer: Mood and Modality. Cambridge University Press. 2001.

[19] L. Polanyi, A. Zaenen: Contextual valence shifters. In Computing Attitude and Affect in Text: Theories and Applications. pp. 110, 2006.

[20] C. Schommer, C. Uhde: Textual Fingerprinting with Texts from Parkin, Bassewitz, and Leander. CoRR abs/0802.2234: (2008).

[21] J. G. Shanahan, Y. Qu, and J. Wiebe: Computing Attitude and Affect in Text: Theory and Applications. Springer, 2006.

[22] A. Siddharthan, S. Teufel: Whose idea was this, and why does it matter? Attributing scientific work to citations. In NAACL-HLT, 2007.

[23] S. Teufel: Argumentative zoning for improved citation indexing. In Computing Attitude and Affect in Text: Theories and Applications, pp. 159169, 2006.

[24] S. Teufel, M. Moens: Discourse level argumentation in scientific articles: human and automatic annotation. Towards Standards and Tools for Discourse Tagging. ACL 1999 Workshop, 1999.

[25] R. Witte, J- Müller (edt.): Text Mining: Wissensgewinnung aus natürlichsprachigen Dokumenten, Interner Bericht 2006-5. Universität Karlsruhe, Fakult at f ür Informatik, Institut für Programmstrukturen und Datenorganisation (IPD), 2006. ISSN 1432-7864.

[26] G. Zhou and J. Su. Named entity recognition using an hmm-based chunk tagger, 2002.